Gene Search

Introduction

Home > Find Genes > Gene Search

Gene Search

Gene Search is a way to search, browse and select genes in IMG. It can be accessed via the second-level menu of **Find Genes**, as shown in Figure 1.

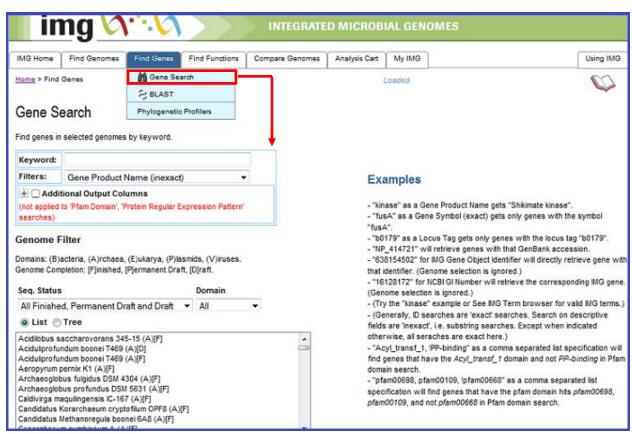


Figure 1: Gene Search.

Keyword Search

User can search for genes containing a keyword in a specific field of the IMG database. A total of 12 search filters are available for selection, as shown in Figure 2.

The default filter is "Gene Product Name". User can also select filters such as "Gene Symbol", "Locus Tag", "GenBank Accession", "NCBI GI Number", "IMG Gene Object ID", "IMG Term and Synonyms", "SEED Product Name/Subsystem", "Is Obsolete Gene" and "Is Pseudo Gene". Further, user can find genes by entering Pfam ID(s) against Pfam Domain or entering a regular expression match against proteins.

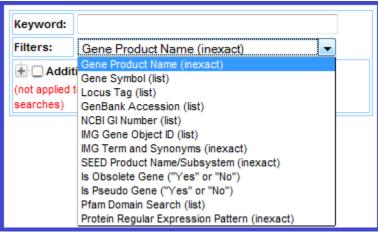


Figure 2: Gene Search Filters.

In addition to search filters, user is also provided the capability to select and display "Additional Output Columns" in result page, as shown in Figure 3:

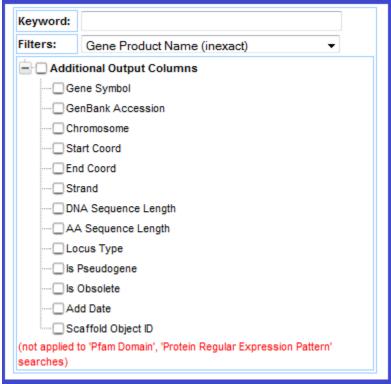


Figure 3: Additional Output Columns for Selection and Display.

These additional display columns are arranged into the following five groups:

- 1. Gene Symbol; GenBank Accession
- 2. Chromosome; Start Coord; End Coord; Strand; DNA Sequence Length; AA Sequence Length; Locus Type
- 3. Is Pseudogene; Is Obsolete
- 4. Add Date
- 5. Scaffold

Genome Filter

The Genome Filter is provided to limit the scope of search. The search, except "NCBI GI Number" and "IMG Gene Object ID" filters, is conducted on the genomes that user has saved through the **Genome Browser**. If user saves nothing, by default, it's all genomes in IMG database. However, user can override the selection via the **Genome Filter** provided in the page.

Since genome selection is ignored in "NCBI GI Number" and "IMG Gene Object ID". The **Genome Filter** section turns hidden when user selects these two filters, and will display again when user selects other filters.

For more information about **Genome Filter**, go to "Home > User Guide > Site Map".

Search Specifications

All searches are case-insensitive. In general, ID or number or symbol searches are 'exact' searches. Searches on descriptive fields are 'inexact', that is, the keyword is used as a word or part of a word.

For example, (i) searching for "fusA" with the filter set to "Gene Symbol" only gets genes with the symbol "fusA"; searching for "bsu02690" with the filter set to "Locus Tag" only gets genes with locus tag "BSU02690"; (ii) searching for "kin" with the filter set to "Product Name" gets "Shikimate kinase," etc.

User can use a percent sign (%) as a wildcard in the middle of a keyword. The results will include any genes with zero or more additional characters at that position. For example, "hydro%ase" will get results with "hydrolase" and "hydrogenase". If user wants only a single character of the keyword to be variable, type an underscore (_) in that position. Searching for "hydro_ase" will get results with "hydrolase," not "hydrogenase."

For regular-expression protein searches, standard regular expression constructs are used against a sequence of amino acid residues. Common constructs include:

- . Matches and single amino acid
- \$ Matches end of sequence.
- ^ Matches beginning of sequence.

- Matches zero or more occurrences of the preceding residue.
 + Matches one or more occurrences of the preceding residue.
 [] Matches any of residues between brackets.
 [^] Matches any of residues not between brackets.